



# Requirements

## Software Specifications

- Ubuntu Server 22.04 LTS:
  - Basic Linux proficiency.
  - Ability to remotely connect via SSH.

## Minimal Secure Cloud Specifications


- Latest NVIDIA GPUs, at least 30xx or RTX A4000 or higher.
  - Option 1: At least 20 GPUs in total, each with a minimum of 12 GB VRAM.
  - Option 2: At least 8 GPUs in total, each with a minimum of 80 GB of VRAM.
  - PCIe 3.0 x16/PCIe 4.0 x8 per GPU or faster. PCIe 4.0 x16 recommended.
  - Demand is highest for SXM 80GB, PCI 80GB, 4090 24GB and L40 or Ada A6000 48GB models. We also require 2 GPU per server at minimum for best rental rates. 8x configurations are the most popular.
- Minimum of 4 Physical CPU Cores Per GPU + 2 for system operations. Prioritize as fast as possible CPU core clock over more cores. For example, a 24 cores CPU clocking at 5.0 GHz is preferred to a 128 cores CPU clocking at 3.0 GHz for a 4 GPU configuration.
- Your RAM should at minimum equals your total vRAM of all GPUs + 12GB for system operations.
  - 1 TB+ of RAM recommended for 8x 80GB vRAM GPU configurations.
- We want an absolute minimum of 1TB+ of NVME space per GPU for each server, ideally 2TB+ (excluding the OS drives). For larger GPU like 80GB, 4TB per GPU is recommended.
  - At least 3,000 MB/s read/write speed is required.
  - We recommend 2 smaller NVME disk in RAID 1 for the operating system (2x 500GB is fine).
  - For data drives, it can be any number of those, but make sure the total is at least 2TB+ per GPU. If several data drives are provided, you need to create a LVM volume for those.

**Ask AI**

- 10gbps Bidirectional Internet Speed as backbone, and minimum of 1gbps symmetrical per server.
- Static Public IP.
  - A single IP can be shared between multiple servers.
  - Access and ability to port forwarding. At least 30 ports forwarded per GPU per server.
- Abide by Tier III+ Datacenter Standards.
  - Robust Uninterruptible Power Supply and backup generator.
  - Internet Service Provider redundancy.
  - 24/7 on-site security and technical staff.
- 400+ GPU Capacity.
- 40GB/s Interconnect between servers.
- Ability to deploy network storage cluster.
- Most **importantly**: the ability to **scale** GPU supply over time.

## PCIe Risers

If your system uses PCIe risers, they **must** have redriver functionality.

 [Edit this page](#)

Previous

« [Maintenance and reliability](#)

Next

[Troubleshooting](#) »

## Docs

[Overview](#)

[Tutorials](#)

[AI APIs](#)

## Community

[Discord](#) 


[Contact us](#) 



**Ask AI**

## More

[Blog](#) 

[GitHub](#) 

Copyright © 2024 RunPod



**Ask AI**